

Syntactic variation and probabilistic indigenization in World Englishes

We report on an ongoing project combining variationist methods (e.g. Tagliamonte 2001) with the Probabilistic Grammar framework (e.g. Bresnan et al. 2007) to study syntactic variation within and across nine varieties of English from around the world: British, New Zealand, Canadian, Irish, Indian, Singaporean, Hong Kong, Philippines, and Jamaican English. Our main interest lies in the extent to which language users' grammatical knowledge differs across these speech communities. In particular, we seek to shed light on questions regarding the extent to which different English varieties share a core grammar, the degree to which individual probabilistic constraints are cross-lectally malleable, and the degree to which certain syntactic alternations exhibit cross-constructional parallelisms. In treating variation as a "core explanandum" (Adger and Trousdale 2007: 274), the project is thus situated at the crossroads of variationist linguistics, usage-based theoretical linguistics, cognitive linguistics, and dialectology.

Our case study explores three well-known patterns of syntactic variation in English, drawing on data from the International Corpus of English (ICE): the genitive alternation (*Mary's speech* versus *the speech of Mary*), the dative alternation (*Tom sent Mary a letter* versus *Tom sent a letter to Mary*), and particle placement (*Tom looked the word up* versus *Tom looked up the word*). In total, our data comprise 8-10k observations of each construction, which were semi-automatically extracted from the ICE corpora: $N = 10594$ genitives; $N = 8549$ datives; $N = 8072$ particle verbs. We carefully circumscribe the variable contexts to identify variants in the corpora, annotating for numerous factors conditioning the choice of constructional variant. Such factors include the semantics of verbs (and particles), as well as the animacy, frequency, definiteness, givenness, and length (in characters) of all construction constituents. Throughout the extraction and annotation process, both automatic (via Perl/Python scripts) and manual coding techniques are employed to maximize efficiency and accuracy. We then use multivariate techniques such as generalized linear mixed-models (Pinheiro and Bates 2000) to investigate variable effects in the conditioning factors that constrain syntactic choices. The effects of such conditioning factors, e.g. the tendency for longer constituents to follow shorter ones, can be seen as stochastic generalizations about language usage, which—according to experimental evidence (Bresnan and Ford 2010)—language users implicitly know about. Thus, we aim to illuminate the variability in the linguistic knowledge that language users with differing English backgrounds implicitly command.

Overall, mixed-effects logistic regression models fit the data quite well for each construction: genitives model, $C = .98$; datives model, $C = .97$; particle verbs model, $C = .92$. We find that the direction of the influence of individual factors is largely consistent across all nine varieties, but that a few cross-varietal differences in the strength of their influence nevertheless emerge in small corners of the data. Such differences are more likely to emerge in outer circle (non-native) varieties, e.g. Hong Kong or Indian English, than inner circle varieties, e.g. Canadian English, but there are few consistent patterns with respect to which factors differ in specific varieties. Interestingly, we also find that some alternations, e.g. particle placement, involve a greater number of varying factors than others, suggesting that different alternation phenomena are not equally amenable to cross-varietal innovation. We speculate that variation in expectation-based processing effects, which are grounded in usage probabilities, could explain the variation exhibited across different varieties. This is supported by recent work showing that the more a constructional alternate is lexically entrenched (i.e. occurs with the same lexical items, Hoffmann 2014) the more likely it is to exhibit cross-varietal variation (AUTHORS forthcoming). Furthermore, an alternation's degree of entrenchment in a given variety correlates inversely with that variety's placement along the path of development in Schneider's (2007) Dynamic Model of nativization. This model predicts that innovations in developing varieties of English typically occur at the syntax-lexicon interface, where new patterns emerge as differences in the habitual associations of constructions with specific lexical items. Results of our study thus reflect the reshaping of stochastic patterns of internal linguistic variation due to shifting usage frequencies among speakers of post-colonial varieties, a process we refer to as 'probabilistic indigenization'. We argue these findings support a probabilistic model of linguistic knowledge which is shaped both by general, higher-level cognitive factors as well as by surface level, community-specific usage norms.

[Complete lists of datasets by variety, conditioning factors in our models, and model summary statistics are provided online at <http://tinyurl.com/oct94j8>]

References

- Adger, D. and G. Trousdale (2007). Variation in English syntax: Theoretical implications. *English Language and Linguistics* 11, 261–278.
- Bresnan, J., A. Cueni, T. Nikitina, and B. Harald (2007). Predicting the dative alternation. In G. Boume, I. Kraemer, and J. Zwarts (Eds.), *Cognitive Foundations of Interpretation*, pp. 69–94. Amsterdam: Royal Netherlands Academy of Science.

- Bresnan, J. and M. Ford (2010). Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1), 168–213.
- Hoffmann, T. (2014). The cognitive evolution of Englishes: The role of constructions in the dynamic model. In S. Buschfeld, T. Hoffmann, M. Huber, and A. Kautzsch (Eds.), *Varieties of English Around the World*, Volume G49, pp. 160–180. Amsterdam: John Benjamins Publishing Company.
- Pinheiro, J. C. and D. M. Bates (2000). *Mixed Effects Models in S and S-PLUS*. New York: Springer.
- Schneider, E. (2007). *Postcolonial English: Varieties Around the World*. New York: Cambridge University Press.
- Tagliamonte, S. (2001). Comparative sociolinguistics. In J. Chambers, P. Trudgill, and N. Schilling-Estes (Eds.), *Handbook of Language Variation and Change*, pp. 729–763. Malden and Oxford: Blackwell.